Xinya Du

Contact

Assistant Professor
Department of Computer Science
The University of Texas at Dallas
800 West Campbell Road, Richardson, TX 75080

Email: xinya.du@utdallas.edu
Website: https://xinyadu.github.io
Google Scholar

Research Interests

Natural Language Processing, Machine Learning, Large Language Models, AI for Science

Education

Cornell University Aug 2016 – Aug 2021
Ph.D. in Computer Science (*M.S.* degree granted in Aug 2019)
Advisor: Claire Cardie

Shanghai Jiao Tong University Sep 2012 – Aug 2016
B.E. in Computer Science and Engineering
Outstanding Graduates Award

Professional Experience

University of Texas at Dallas, Richardson, TX Aug 2022 – Present
Assistant Professor in Computer Science

University of Illinois at Urbana-Champaign, Champaign, IL Sep 2021 – Aug 2022
Postdoctoral Research Associate, with Prof. Heng Ji

Industrial Experience

Google AI, Mountain View, CA May 2020 – Aug 2020
Research Intern

Allen Institute for Artificial Intelligence, Seattle, WA Sep 2018 – Dec 2018
Research Intern

Microsoft Research, Redmond, WA May 2018 – Aug 2018
Research Intern

Selected Awards& Honors

NSF CAREER award National Science Foundation, 2024
AAAI New Faculty Highlights AAAI, 2024
Cisco Faculty Research Award Cisco, 2024
Amazon Research Award Amazon, 2023
Spotlight Rising Star in Data Science University of Chicago, 2021
Top 100 New Stars in Artificial Intelligence Baidu Scholar, 2020
Most Influential ACL Papers (15 each year) Paperdigest, 2017
National Scholarship (Top 1% students nationwide) SJTU, 2013

Publications

Note: † indicates that I am a co-leading author. * indicates equal contributions.
You can also find my publication list on [DBLP] and [Google Scholar] pages.

[1] **IQA-EVAL: Automatic Evaluation of Human-Model Interactive Question Answering** Ruosen Li, Ruochen Li, Barry Wang, **Xinya Du**
In Annual Conference on Neural Information Processing Systems (NeurIPS), 2024.

[2] **FaithScore: Fine-grained Evaluations of Hallucinations in Large Vision-Language Models**
Liqiang Jing, Ruosen Li, Yunmo Chen, **Xinya Du**
In Findings of the Association for Computational Linguistics: (EMNLP), 2024.

[3] **MEQA: A Benchmark for Multi-hop Event-centric Question Answering with Explanations**
Ruosen Li, Zimu Wang, Son Quoc Tran, Lei Xia, **Xinya Du**
In Annual Conference on Neural Information Processing Systems (NeurIPS), 2024.

[4] **Document-level Causal Relation Extraction with Knowledge-guided Binary Question Answering**
Zimu Wang, Lei Xia, Wei Wang, **Xinya Du**
In Findings of the Association for Computational Linguistics: (EMNLP), 2024.

[5] **QAEvent: Event Extraction as Question-Answer Pairs Generation**
Milind Choudhary, **Xinya Du**
In Findings of the Association for Computational Linguistics: (EACL), 2024.

[6] **PRD: Peer Rank and Discussion Improve Large Language Model-based Evaluations**
Ruosen Li, Teerth Patel, **Xinya Du**
In Transactions on Machine Learning Research (TMLR), 2024.

[7] **Large Language Models for Automated Open-domain Scientific Hypotheses Discovery**
Zonglin Yang, **Xinya Du**†, Junxian Li, Jie Zheng, Soujanya Poria, Erik Cambria
In Findings of the Association for Computational Linguistics: (ACL), 2024.
Best poster award in AI4Science workshop, 2024 (1/200).

[8] **Making Natural Language Reasoning Explainable and Faithful**
Xinya Du
In AAAI Conference on Artificial Intelligence (AAAI), 2024.

[9] **Language Models as Inductive Reasoners**
Zonglin Yang, Li Dong, **Xinya Du**†, Hao Cheng, Erik Cambria, Xiaodong Liu, Jianfeng Gao, Furu Wei
In Conference of the European Chapter of the Association for Computational Linguistics (EACL), 2024.

[10] **Leveraging Structured Information for Explainable Multi-hop Question Answering and Reasoning**
Ruosen Li, **Xinya Du**
In Findings of the Association for Computational Linguistics: (EMNLP), 2023.

[11] **Process of Elimination for Multiple Choice Reasoning**
Chenkai Ma, **Xinya Du**
In Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.

[12] **Probing Representations for Document-level Event Extraction**
Barry Wang, **Xinya Du**, Claire Cardie
In Findings of the Association for Computational Linguistics: (EMNLP), 2023.

[13] **Zero-Shot Classification by Logical Reasoning on Natural Language Explanations**
Chi Han, Hengzhi Pei, **Xinya Du**, Heng Ji
In Findings of the Association for Computational Linguistics: (ACL), 2023.

[14] **Toward Consistent and Informative Event-Event Temporal Relation Extraction**
Xiaomeng Jin, Haoyang Wen, **Xinya Du**, Heng Ji
In MATCHING at Annual Meeting of the Association for Computational Linguistics (ACL), 2023.

[15] **End-to-end Case-Based Reasoning for Commonsense Knowledge Base Completion**
Zonglin Yang, **Xinya Du**†, Erik Cambria, Claire Cardie
In Conference of the European Chapter of the Association for Computational Linguistics (EACL), 2023.

[16] **Logical Entity Representation in Knowledge-Graphs for Differentiable Rule Learning**
Chi Han, Qizheng He, Charles Yu, **Xinya Du**, Hanghang Tong, Heng Ji
In International Conference on Learning Representations (ICLR), 2023.

[17] **RESIN-11: Schema-guided Event Prediction for 11 Newsworthy Scenarios**
Xinya Du, Zixuan Zhang, Sha Li, Heng Ji and the RESIN team
In Conference of the North American Chapter of the Association for Computational Linguistics (NAACL): System Demonstrations, 2022.
Top ranking system in DARPA KAIROS evaluation.

[18] **Retrieval-Augmented Generative Question Answering for Event Argument Extraction**
Xinya Du and Heng Ji
In Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022.

[19] **Dynamic Global Memory for Document-level Argument Extraction**
Xinya Du, Sha Li, and Heng Ji
In Annual Meeting of the Association for Computational Linguistics (ACL), 2022.

[20] **Automatic Error Analysis for Document-level Information Extraction**
Aliva Das*, **Xinya Du***, Barry Wang*, Kejian Shi, Jiayuan Gu, Thomas Porter, Claire Cardie
In Annual Meeting of the Association for Computational Linguistics (ACL), 2022.

[21] **Template Filling with Generative Transformers**
Xinya Du, Alexander M. Rush, and Claire Cardie
In Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2021.

[22] **GRIT: Generative Role-filler Transformers for Document-level Event Entity Extraction**
Xinya Du, Alexander M. Rush, and Claire Cardie
In Conference of the European Chapter of the Association for Computational Linguistics (EACL), 2021.

[23] **Few-shot Intent Classification and Slot Filling with Retrieved Examples**
Dian Yu, Luheng He, Yuan Zhang, **Xinya Du**, Panupong Pasupat and Qi Li

In Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2021.

[24] **QA-Driven Zero-shot Slot Filling with Weak Supervision Pretraining**
Xinya Du, Luheng He, Qi Li, Dian Yu, Panupong Pasupat and Yuan Zhang
In Annual Meeting of the Association for Computational Linguistics (ACL), 2021.

[25] **Event Extraction by Answering (Almost) Natural Questions**
Xinya Du and Claire Cardie
In Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.
Top 1% most cited articles published in Computer Science in 2020.

[26] **Improving Event Duration Prediction via Time-aware Pre-training**
Zonglin Yang, **Xinya Du**, Alexander M. Rush and Claire Cardie
In Findings of the Association for Computational Linguistics: (EMNLP), 2020.

[27] **Document-Level Event Role Filler Extraction using Multi-Granularity Contextualized Encoding**
Xinya Du and Claire Cardie
In Annual Meeting of the Association for Computational Linguistics (ACL), 2020.

[28] **Leveraging Structured Metadata for Improving Question Answering on the Web**
Xinya Du, Adam Fourney, Robert Sim, Claire Cardie, Paul Bennett and Ahmed Hassan Awadallah
In Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (AACL/IJCNLP), 2020.

[29] **Be Consistent! Improving Procedural Text Comprehension using Label Consistency**
Xinya Du, Bhavana Dalvi, Niket Tandon, Antoine Bosselut, Wen-tau Yih, Peter Clark, Claire Cardie
In Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019.

[30] **Harvesting Paragraph-Level Question-Answer Pairs from Wikipedia**
Xinya Du and Claire Cardie
In Annual Meeting of the Association for Computational Linguistics (ACL), 2018.
Top 1% most cited articles published in Computer Science in 2018.

[31] **Identifying Where to Focus in Reading Comprehension for Neural Question Generation**
Xinya Du and Claire Cardie
In Conference on Empirical Methods in Natural Language Processing (EMNLP), 2017.

[32] **Learning to Ask: Neural Question Generation for Reading Comprehension**
Xinya Du, Junru Shao and Claire Cardie
In Annual Meeting of the Association for Computational Linguistics (ACL), 2017.
Featured in New Scientist and TechRepublic **[Link]**
Top 0.1% most cited articles published in Computer Science in 2017

[33] **Cornell Belief and Sentiment System at TAC 2016**
Vlad Niculae, Kai Sun, Xilun Chen, Yao Cheng, **Xinya Du**, Esin Durmus, Arzoo Katiyar, Claire Cardie
In Text Analysis Conference (TAC), 2016.

Grants

NSF CAREER: Learning to Extract Consistent Event Graphs from Long and Complex Documents (PI)
Funding Source: Natural Science Foundation.
Period: May 2024 – present.
Amount Awarded: $561,219. Amount to me: $561,219.

Process-guided Fine-tuning for Answering Complex Questions (PI)
Funding Source: Amazon Research Award.
Period: Jan 2024 – present.
Amount Awarded: $110,000 ($40,000 AWS credits). Amount to me: $110,000.

Amazon Trusted AI Challenge Grant (Co-PI)
Funding Source: Amazon.
Period: Sep 2024 – present.
Amount Awarded: $250,000. Amount to me: $125,000.

FAIGen: Faithful LLM Generation with Scientific Principles-guided Learning (PI)
Funding Source: Cisco Faculty Research Award.
Period Jan 2025 – present.
Amount Awarded: $50,000. Amount to me: $50,000.

Faithfulness in Large Vision Language Models (PI)
Funding Source: OpenAI Researcher Access Program (jointly awarded with PhD student).
Period: Sep 2024 – present.
Amount Awarded: $4,000. Amount to me: $4,000.

Undergraduate Research Apprenticeship Award (URAP) (PI)
Funding Source: University of Texas at Dallas.
Period: Summer 2023, 2024.
Amount Awarded: $9,000.

Teaching Experience

Introduction to Machine Learning, UT Dallas, Spring 2024

Natural Language Processing, UT Dallas, Fall 2023, 2024

Deep Learning for Natural Language Processing, UT Dallas, Spring 2023
New course developed by me (overall score of 4.65/5.0) [Link]

Natural Language Processing, UT Dallas, Fall 2022

Natural Language Processing, Cornell University, Fall 2019
Teaching Assistant for Prof. Claire Cardie.

Natural Language Processing, Cornell University, Spring 2019
Teaching Assistant for Prof. Yoav Artzi.

Software Engineering, Cornell University, Spring 17, Spring 18
Teaching Assistant for Prof. William Arms.

Introduction to Computing Using Python, Cornell University, Fall 2016
Teaching Assistant for Prof. Walker White.

Mentoring Experience

PhD students
Ruosen Li (2022–present, UT Dallas PhD student)
Topic: Large language model peer evaluations.

Publications: TMLR 2024, NeurIPS 2024, EMNLP (Findings) 2023 2024.

Liqiang Jing (2023–present, UT Dallas PhD student)
Topic: Large language model peer evaluations.
Publications: EMNLP (Findings) 2024.
Award: OpenAI Researcher Access Program.

Ruochen Li (2023–present, UT Dallas PhD student)
Topic: Large language model peer evaluations.
Publications: NeurIPS 2024.

Guiming Chen (2024–present, UT Dallas PhD student)
Topic: Large language model peer evaluations.

Master Students

Milind Choudhary (2023, UT Dallas Master student) → UT Dallas PhD
Topic: Event Extraction as Question-Answer Pairs Generation
Publications: EACL 2024 (Findings).

Zimu Wang (2023–present, UT Dallas Visiting student)
Topic: Temporal relation extraction.
Publications: EMNLP 2024 (Findings), NeurIPS 2024.

Son Tran (2023, UT Dallas Visiting student) → Cornell CS PhD
Topic: Question Answering.
Publications: NeurIPS 2024.

Chenkai Ma (2023, UT Dallas Visiting student)
Topic: Multiple Choice Questions Reasoning.
Publications: EMNLP 2023.

Zonglin Yang (2020–2024, Cornell CS MEng student → NTU PhD)
Topic: Commonsense and Case-based Reasoning for NLP.
Publications: EMNLP (Findings) 2020, EACL 2023, ACL 2024 (Findings).

Barry Wang (2021–2024, Cornell CS undergraduate student) → CMU PhD
Topic: Automatic Error Analysis for Information Extraction.
Publications: ACL 2022, SciNLP 2022, EMNLP 2023 (Findings), NeurIPS 2024.

Undergraduate/High School Students

Teerth Patel (UT Dallas BS student, 2023–present)
Topic: Large language model peer evaluations.
Publications: TMLR 2024.

Lei Xia (2023–2024, UT Dallas Visiting student)
Topic: Multi-hop Question Answering.
Publications: NeurIPS 2024.

Minhao Zou (2024, UT Dallas Visiting student)
Topic: LLM for Scientific Discovery.

Arjun Junghare (2024, UT Dallas CS undergraduate student)
Topic: Image Generation with Large Language Models.

Jaden Nunes (Summer 2023, DFW Local K-12 student)
Topic: Event Extraction as Question-Answer Pairs Generation.

Rishab Bhattacharya (Summer 2023, DFW Local K-12 student)
Topic: Event Extraction as Question-Answer Pairs Generation.

Shreyas Kumar (Summer 2023, DFW Local K-12 student)
Topic: Event Extraction as Question-Answer Pairs Generation.

Rishi Malhotra (Spring 2021, Cornell CS undergraduate student → Microsoft)
Topic: Applying Neural Document-level IE Model to the Scientific Domain.

Aliva Das (2021–2022, Cornell CS undergraduate student → Amazon)
Topic: Automatic Error Analysis for Information Extraction.
Publications: ACL 2022, SciNLP 2021.

Maitreyi Chatterjee (Spring 2021, Cornell CS undergraduate student → LinkedIn)
Topic: Applying Neural Document-level IE Model to the Scientific Domain.

Professional Services

Chairing:

ACL Rolling Review (ARR) 2024

Conference on Empirical Methods in Natural Language Processing (EMNLP) 2024 Area char

Annual Meeting of the Association for Computational Linguistics (ACL) 2023 Area char

International Conference on Computational Linguistics (COLING) 2024, 2025 Senior Area Chair

Conference on Empirical Methods in Natural Language Processing (EMNLP) Demo Track 2024 Area char

Conference of the North American Chapter of the Association for Computational Linguistics (NAACL) 2024 Website chair

Annual Meeting of the Association for Computational Linguistics (ACL) 2024 Session Chair

Seminar/Conference/Workshop Organizing:

AI4Research: Towards a Unified Knowledge-grounded Scientific Research Lifecycle. To Appear AAAI 2025. Organizing team: Qingyun Wang, Wenpeng Yin, Lifu Huang, Yi R. Fung, Xinya Du, Carl Edwards, Tom Hope.

Journal Reviewer:

IEEE Transactions on Knowledge and Data Engineering (TKDE)

IEEE Transactions on Audio, Speech and Language Processing (TASLP)

IEEE Transactions on Neural Networks and Learning Systems (TNNLS)

ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)

ACM Transactions on Knowledge Discovery from Data (TKDD)

Computational Linguistics (CL)

Knowledge and Information Systems (KAIS)

AI Communication

Information Processing and Management (IPM)

Conference Committee Member:

Annual Meeting of the Association for Computational Linguistics (ACL)

Annual Conference on Neural Information Processing Systems (NeurIPS)

International Conference on Learning Representations (ICLR)

International Conference on Artificial Intelligence and Statistics (AISTATS)

Conference on Empirical Methods in Natural Language Processing (EMNLP)

Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)

Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (AACL/IJCNLP)

International Joint Conference on Artificial Intelligence (IJCAI)

AAAI Conference on Artificial Intelligence (AAAI)

Natural Language Processing and Chinese Computing (NLPCC)

Conference on Computational Natural Language Learning (CoNLL)

Workshop on Noisy User-generated Text (W-NUT)

Workshop on Machine Reading for Question Answering (MRQA)

Joint Conference on Lexical and Computational Semantics

PhD/MS Committee Member:

PhD: Xiangci Li, Jishnu Jaykumar Padalunkal, Basel Abdeen, Yibo Hu.

MS: Wooseong Yang, Shubham Patel

Other Activities:

Faculty advisor and judge for Association for Computing Machinery Symposium (ACM). UT Dallas. 2023-2024.

Faculty Advisor for Women Who Compute (WWC). UT Dallas. 2024.

Faculty Advisor for Girls Who Code (GWC). UT Dallas. 2024.

Faculty Advisor for Society of Asian Scientists and Engineers (SASE). UT Dallas. 2024

Research Mentor for RIDE (Research, Inquiry, Design Experience) Project. UT Dallas. 2024.

Member of PhD Admission Committee. UT Dallas. 2022, 2023, 2024

Site host of North American Computational Linguistics Olympiad (NACLO). 2023, 2024.

Member of Cornell CS Department PhD Admission Committee. 2021.

Volunteer for Cornell CS Department PhD Visit Day. 2019, 2020, 2021.

Student Volunteer for ACL 2017, ACL 2018, EMNLP 2017.

Recent Talks

Synergizing Knowledge and Large Language Models
University of Illinois Urbana-Champaign, Data Mining Group Seminar, Oct 2024.

Synergy between Large Language Model and Knowledge
University of North Texas, Oct 2024.

Synergizing Knowledge and Large Language Models
University of Massachusetts–Amherst, Machine Learning & Friends Lunch, Sep 2024.

Faculty Round Table Talk
UT Dallas Hobson Wildenthal Honors College, Aug 2024.

Large Language Models: Knowledge, Reasoning and Factuality
Samsung Electronics America, Mar 2024.

Open-ended Evaluations of Foundational Models: Alignment and Faithfulness
Shanghai Jiao Tong University, Computer Science Department Seminar, Dec 2023.
Fudan University, Computer Science Department Seminar, Dec 2023.

ChatGPT: Fact vs. Fiction
UT Dallas, Forum sponsored by The Dallas Morning News, May 2023.

Towards More Intelligent Extraction of Information from Documents
Allen Institute for Artificial Intelligence, Feb 2022
Simon Fraser University, Feb 2022

Rensselaer Polytechnic Institute, Feb 2022
University of California, Merced, Feb 2022

Towards More Intelligent Extraction of Information from Documents
UIUC, Siebel School of Computing and Data Science Speaker Series, Feb 2022

Towards More Informed Extraction of Events from Documents
University of Chicago, Rising Stars in Data Science Workshop, Jan 2021.
Tencent AI Research America, Nov 2020.

Event Extraction by Answering (Almost) Natural Question
UIUC, Information Extraction and Knowledge Acquisition Class, Sep 2020

LwLL: Progress on the NLP Front
Cornell University, DARPA site visit, Apr 2020.

Harvesting Paragraph-Level Question-Answer Pairs from Wikipedia
56th Annual Meeting of the Association for Computational Linguistics, July 2018.